ELSEVIER

Remote Sensing of Environment 76 (2001) 418–422

Remote Sensing of Environment

www.elsevier.com/locate/rse

Short communication

Thematic accuracy of MRLC land cover for the eastern United States

Limin Yang[a], Stephen V. Stehman[b], Jonathan H. Smith[c,*], James D. Wickham[c]

[a]Raytheon-ITSS, EROS Data Center, Sioux Falls, SD 57198, USA
[b]SUNY-ESF, 320 Bray Hall, Syracuse, NY 13210, USA
[c]US Environmental Protection Agency (MD-56), Research Triangle Park, NC 27711, USA

Received 12 June 2000; accepted 29 December 2001

Abstract

One objective of the MultiResolution Land Characteristics (MRLC) consortium is to map general land-cover categories for the conterminous United States using Landsat Thematic Mapper (TM) data. Land-cover mapping and classification accuracy assessment are complete for the eastern United States. The accuracy assessment was based on photo-interpreted reference data obtained from a stratified probability sample of pixels. Agreement was defined as a match between primary or alternate reference land-cover labels assigned to each sample pixel and the mode (most common class) of the map's land-cover labels within a 3 × 3-pixel neighborhood surrounding the sampled point. At 30-m resolution, overall accuracy was 59.7% at an Anderson Level II thematic detail, and 80.5% at Anderson Level I. © 2001 Elsevier Science Inc. All rights reserved.

1. Introduction

Information on land cover (and use) supports a broad range of studies concerned with characteristics of the earth's surface, especially environmental studies and land use planning. In the United States, Landsat satellite data are used routinely to map land cover to serve these purposes. Yet, despite the numerous Landsat-based land-cover mapping studies that have been undertaken in this country, the geographic scope of these studies typically has been limited to state boundaries or even more local jurisdictions.

Recognizing a need for remote sensor and derivative data of national scope, several federal agencies (US Geological Survey, Environmental Protection Agency, National Oceanographic and Atmospheric Administration, US Fish and Wildlife Service, and US Forest Service) formed the MultiResolution Land Characteristics Consortium (MRLC) to purchase and process Landsat Thematic Mapper (TM) data (Loveland & Shaw, 1996). One of MRLC's activities has been the creation of the National Land Cover Data (NLCD) set, a land-cover map of the conterminous United States derived from classified Landsat TM data. The data set is complete and is available through the USGS Earth Resources Observation Systems (EROS) Data Center's homepage at http://edc.usgs.gov/programs/lccp/natllandcover.html. This data set is the first known national-scope and consistently classified land-cover data set for the country. The USGS's previous land-use and land-cover data (commonly referred to as LUDA) were based on early 1970s National High Altitude Photography (NHAP) using several photointerpreters, and some areas were never completed (Fegeas, Claire, Guptill, Anderson, & Hallam, 1983).

Land-cover classification was accomplished through unsupervised clustering (Kelly & White, 1993) of either leaves-on or leaves-off images, the majority of which were acquired during the early 1990s. Resulting spectral clusters (Vogelmann, Seevers, & Oimoen, 1997) were resolved into 1 of 21 thematic classes using logical modeling and several ancillary data sources (e.g., census, slope/aspect/elevation). The 21 thematic classes resemble the well-established Anderson land use/cover classification system (Anderson, Hardy, Roach, & Witmer, 1976). Details of the classification process are discussed in Vogelmann, Sohl, Campbell, and Shaw (1998) and Vogelmann, Sohl, and Howard (1998). Positional accuracies of the geometrically corrected Landsat TM images utilized in the classification process were on the order of ±1 pixel.

Though land-cover mapping of the conterminous United States is complete, accuracy assessments are continuing, with the eastern United States having been completed

* Corresponding author. Tel.: +1-919-541-5029.
E-mail address: smith.jonathanh@epa.gov (J.H. Smith).

0034-4257/01/$ – see front matter © 2001 Elsevier Science Inc. All rights reserved.
PII: S0034-4257(01)00187-0

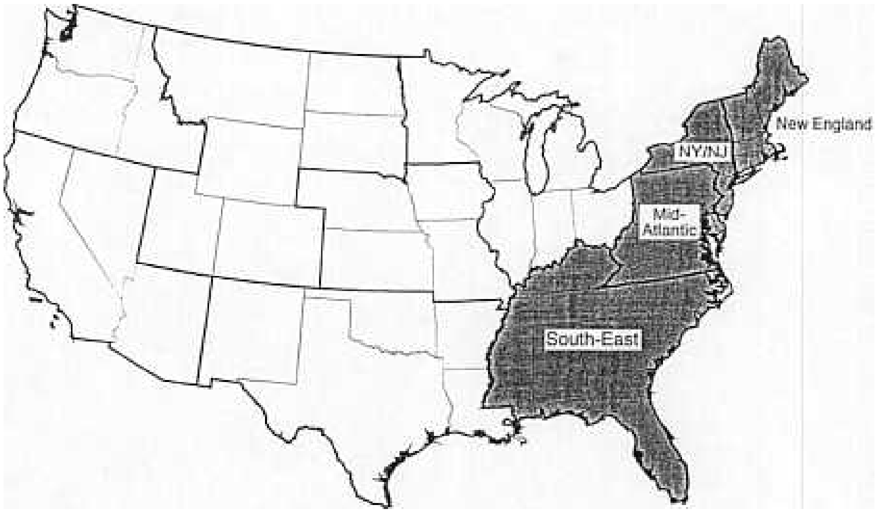

Fig. Accuracy assessed area.

(Fig. 1). This paper provides a condensed report of the land-cover accuracy assessment results for this region.

2. Methods

The sampling design incorporated three layers of stratification and a two-stage cluster sampling protocol. Each mapping region (New England, New York/New Jersey, mid-Atlantic, and Southeast) constituted a stratum and was sampled independently. Within each mapping region, geographic strata were created using $15' \times 15'$ or $30' \times 30'$ grid cells, depending on the size of the region. Primary sampling units (PSU) defined by nonoverlapping, interior regions of aerial photographs were then delineated within these strata. These PSUs partition each region into nearly equal area units. A single PSU was selected from each geographic stratum, with all PSUs having an equal probability of being selected. This first-stage sampling protocol based on the geographic strata was designed to enhance the geographic spread of the sample. The pixels selected within the first-stage PSUs were then stratified by mapped land-cover class, and a simple random sample of pixels was selected independently for each land-cover class.

The New York/New Jersey sampling design differed slightly from the others. In this region, the land-cover class stratification was implemented only for seven rare classes, and these sample data were then combined with pixels obtained from a separate, general design not stratified by land-cover class (see Zhu, Yang, Stehman, & Czaplewski, 2000).

Within each region, the sampling design was equal probability for pixels within a land-cover class. Pixels from different land-cover classes had different probabilities of being included in the sample. Within each mapping region, stratified sampling formulas were applied to estimate the error matrix cell proportions (Stehman & Czaplewski, 1998, pp. 338–340).

To obtain the reference classification, each sample (pixel) of unknown land cover was located on the air photo by context using a drape of the sample point over a Landsat TM composite image, and photointerpreting 1:40,000-scale National Aerial Photography Program (NAPP) black-and-white or color infrared film paper prints. The majority of these photographs were acquired during the period 1989–1993. During the assessment, procedures were specified to insure that the photointerpreters remained unaware of the Landsat TM classification results. Reference information (Table 1) collected by the photointerpreters to help interpret agreement statistics included primary and alternate land-cover labels,, a land-cover heterogeneity score based on the

Table 1
Attribute information collected for interpreting agreement between map and reference data

Information from reference source
(1) Primary reference label: label-cover label thought to be most correct by photointerpreter.
(2) Alternate reference label: land-cover label that might also be considered correct given information in the photograph. An alternate reference label was not provided if the photointerpreter's primary reference label was believed unambiguously evident.
(3) Photointerpreter confidence: nominal ranking of photointerpreter's certainty in identifying reference land-cover label. Rank values range from 1 (*not confident*) to 4 (*very confident*).
(4) Relative location: nominal score for location of sample point (pixel): 1 = on the edge between two land-cover classes; 2 = homogeneous area of land cover; and 3 = heterogeneous area of land cover.
5. Date: day, month, and year of photo acquisition.
6. Notes: entries for any other factors that may affect photointerpretation (e.g., temporal change).
Information from map source
(1) Center: land-cover label of pixel selected as a sample.
(2) Mode: most frequent land-cover label in a 3 × 3-pixel neighborhood surrounding selected sample point.
(3) Any: list of all land-cover labels in a 3 × 3-pixel neighborhood surrounding selected sample point.

complexity of the landscape in the vicinity of the point, and a confidence rating of the photointerpreted land-cover label (Zhu, Yang, Stehman, & Czaplewski, 1999). In this article, we define agreement as a match between the primary or alternate reference land-cover label of the sampled pixel and a mode land-cover label in a 3×3-pixel window surrounding the sample pixel. Use of both the primary and alternate land cover labels in the matching protocol is analogous to the RIGHT operator developed by Gopal and Woodcock (1994). The mode refers to a land cover class that is most prevalent in the 3×3 window. If two or more classes qualify as a mode, then a correct classification would occur if either the primary or secondary reference label agreed with one of the mode classes. All sample data were included in the analyses regardless of photointerpreted land-cover label confidence rating or land-cover heterogeneity scores.

The reported estimates represent an intermediate position in the range of possible accuracy results available from using different definitions of agreement. Estimates based on the primary reference label of the center pixel and using all samples regardless of photointerpreter confidence and relative location scores represent the lower end of the range of results. Such estimates may reflect a 'conservative bias' (Verbyla & Hammond, 1995) because of confounding of true classification error with errors attributable to misregis-

Table 2
Error matrices at Anderson Level II (a) and I (b)

(a) Error matrix at Anderson Level II

	11	21	22	23	31	32	33	41	42	43	51	81	82	85	91	92	Total	*C*	Sample size
11	**55.49**	0.14	0.00	0.16	0.12	0.00	0.56	0.00	0.48	0.00	0.00	0.10	0.06	0.07	0.02	0.94	58.14	0.05	374
21	0.04	**18.18**	1.27	2.55	0.26	0.00	0.27	0.97	1.02	1.34	0.13	0.38	0.28	3.01	0.13	0.19	30.02	0.39	350
22	0.01	1.78	**2.71**	1.50	0.09	0.00	0.03	0.04	0.14	0.04	0.00	0.04	0.03	0.27	0.02	0.02	6.72	0.60	308
23	0.37	1.97	0.09	**7.22**	0.97	0.02	0.15	0.26	0.17	1.90	0.02	0.21	0.35	1.15	0.07	0.18	15.10	0.52	320
31	0.60	1.72	0.00	0.61	**24.67**	0.60	0.01	0.01	0.11	0.05	0.02	7.18	3.59	9.06	0.00	4.21	52.44	0.53	300
32	0.08	0.11	0.00	0.27	0.45	**2.43**	1.97	1.23	0.01	0.20	0.00	1.42	0.06	0.09	0.05	0.02	8.39	0.71	278
33	0.02	0.30	0.00	0.04	0.07	0.00	**9.47**	1.20	2.04	2.22	0.18	1.15	0.60	1.96	1.43	0.36	21.04	0.55	295
41	0.57	3.77	0.09	0.66	0.30	1.36	10.02	**123.14**	8.97	29.56	2.13	9.98	7.19	5.07	3.07	1.01	206.89	0.40	761
42	0.40	0.36	0.21	1.94	2.24	0.00	2.37	6.65	**87.81**	46.48	1.66	2.56	0.32	1.64	10.92	1.31	166.87	0.47	367
43	0.70	1.78	0.00	0.09	0.63	0.00	5.47	15.18	17.07	**109.93**	1.11	3.19	2.39	3.34	2.28	0.19	163.35	0.33	469
51	0.00	0.00	0.00	0.00	0.01	0.00	0.03	0.00	0.01	0.00	**0.17**	0.00	0.00	0.00	0.00	0.00	0.23	0.25	58
81	0.00	1.83	0.18	0.83	0.04	0.00	1.15	4.19	1.53	1.98	0.04	**18.46**	23.00	2.20	0.06	0.02	55.51	0.67	354
82	0.55	1.06	0.00	0.78	0.79	0.11	1.54	2.53	1.86	1.96	0.23	5.48	**21.36**	3.55	0.43	0.96	43.19	0.51	429
85	0.02	1.47	0.74	0.49	0.02	0.00	0.10	0.18	1.01	0.31	0.00	1.32	0.61	**9.87**	0.15	0.01	16.30	0.39	287
91	1.21	0.17	0.10	0.07	0.66	0.02	4.39	5.20	4.19	8.65	0.17	1.30	1.41	0.64	**49.29**	3.18	80.65	0.39	306
92	1.52	0.10	0.82	0.81	1.65	0.00	0.04	0.30	0.27	4.86	0.08	1.60	0.24	0.86	5.12	**56.86**	75.13	0.24	285
																	597.06		
Total	61.58	34.74	6.22	18.02	32.97	4.54	37.57	161.08	126.69	209.48	5.94	54.37	61.49	42.78	73.04	69.46	999.97		
O	0.10	0.45	0.56	0.60	0.25	0.46	0.75	0.24	0.31	0.48	0.97	0.66	0.65	0.77	0.33	0.18			
Sample size	416	467	155	372	180	123	287	690	503	561	127	386	460	288	222	273			5511

(b) Error matrix at Anderson Level I

		20s	30s	40s		80s	90s	Total	*C*	Sample size
11	**51.51**	0.30	0.68	0.48	0.00	0.16	0.96	54.09	0.05	
20s	1.14	**39.53**	1.54	5.32	0.07	5.31	0.62	53.53	0.26	
30s	0.68	1.77	**40.29**	5.21	0.19	24.43	4.83	77.40	0.48	
40s	2.02	9.18	18.35	**467.75**	4.67	34.30	14.31	550.58	0.15	
51	0.00	0.00	0.03	0.01	**0.16**	0.00	0.00	0.20	0.25	
80s	0.84	6.72	3.62	11.80	0.48	**85.59**	1.50	110.55	0.23	
90s	2.72	2.12	5.81	16.02	0.20	6.83	**119.92**	153.62	0.22	
								804.75		
Total	58.91	59.62	70.32	506.59	5.77	156.62	142.14	999.97		
O	0.13	0.34	0.43	0.08	0.97	0.45	0.16			
Sample size	408	994	592	1757	132	1138	490			5511

Reference land-cover labels form the columns and map land-cover labels form the rows of the error matrices. The values in the error matrices are the estimated cell area proportions times 1000 (e.g., 55.49 is 0.05549) rounded off to two digits beyond the decimal point. *O* is omission error (1 minus main diagonal entry divided by the column total). *C* is commission error (1 minus main diagonal entry divided by the row total). Land cover classes are: water (11), low density residential (21), high density residential (22), commercial/industrial/transportation (23), bare rock/sand/clay (31), mining (32), transitional (33), deciduous forest (41), evergreen forest (42), mixed forest (43), shrubland (51), hay and pasture (81), cropland (82), urban grass (85), woody wetland (91), and emergent (herbaceous) wetland (92).

tration or inability to correctly photointerpret a point. Conversely, the upper end of the range of accuracy results would occur when estimates are based on defining agreement as a match between either the primary or alternate reference label with any of the 3 × 3 neighborhood map land-cover labels and using only sample pixels possessing high photointerpreter confidence scores (e.g., scores 3 and 4) and homogeneous land cover (e.g., relative location score of 2). These estimates likely have an 'optimistic bias' (Hammond & Verbyla, 1996) because the sample data are restricted to areas where land cover is more easily identified and a more inclusive definition of agreement is applied.

Agreement results for the four mapping regions were combined by weighting the error matrix cell proportions by the proportion of area in the eastern United States represented by each region. The cell proportions of the resulting error matrix for the eastern United States have been expressed as a proportion of area times 1000. For example, a value of 18.18 reported for the cell in row x and column y indicates that about 1.8% of the total area was estimated to have a map label x and a reference label y. Sample pixels for which a reference classification was not obtained due to missing photographs, clouds being present on the photographs or other problems were treated as missing at random and accounted for less than 7% of the sample points.

3. Results

Overall accuracy for the eastern United States was 59.7% at Anderson Level II and 80.5% at Anderson Level I (Table 2a and b). As expected, a significant source of disagreement between map and reference land-cover labels (approximately 20%) was between classes that aggregate into a single Anderson Level I class. Another significant source of disagreement was between the forest and agricultural classes, which accounted for approximately 5% of the estimated area misclassified. This disagreement may partially result from the photointerpreters being uncertain as to the correct location of sample points on the photographs, especially at those points located near patch boundaries and not due to thematic misclassifications. A third significant source of disagreement was between the forested wetland class (91) and the upland forest classes (41, 42, 43). If confusion between these two sets was not considered disagreement, accuracies at both Levels improved by about 3%.

At Anderson Level II, commission errors tended to be between 30% and 60%. Notably high commission errors occurred for Class 32 (mining), Class 81 (hay/pasture) and Class 22 (high density residential). Class 22 was most often confused with the other urban classes, 21 and 23 (low density residential and commercial/industrial/transportation), while Class 81 was most often confused with Class 82 (row crops). Both of these problems were removed at Anderson Level I. This was not the case, however, with Class 32, which was most often confused with Class 33 (transitional), in addition to Classes 81 and 41 (deciduous forest).

Omission errors showed much more variation among the classes. Classes 33, 51 (shrubland), 81, 82, and 85 (urban grasses) all had omission errors greater than 60%. In each case, the highest sources of confusion were within the Level I classifications and with the upland forest classes. Class 51 was an anomaly because it was mapped only in New England, and the high commission error may reflect a difference in interpretation of the class definition between photointerpreters and map creators.

At Anderson Level I, class-specific commission errors were approximately 25% or lower except for the 30s class. Omission errors were low for Classes 11, 40s (upland forest), and 90s (wetlands), but exceeded 33% for Classes 20s (urban), 30s (barren), and 80s (agriculture). The anomalous class 51 results persisted at Level I because no other 50s classes were present to combine with class 51.

4. Discussion

Several rules for defining agreement between map and reference data may be applied given the information collected. Subsequent description of map accuracy for the eastern United States will include error matrices based on different definitions of agreement and different subsets of the data (e.g., subsets defined by photointerpreter confidence ratings as in Zhu et al., 2000). Comparing results across a range of agreement protocols and data subsets permits evaluation of the reference data quality and a more thorough investigation of thematic map accuracy (Congalton & Green, 1993; Khorram et al., 1999). Analyses exploring sources of classification error and their effects on accuracy results are in progress (Yang, Stehman, Wickham, Smith, & Van Driel, 2000). Accuracy assessment for the remainder of the conterminous United States is ongoing, and detailed accuracy results for the conterminous United States will be reported when complete.

Acknowledgments

This work has been supported by the US Environmental Protection Agency and the US Geological Survey. The work has been reviewed by both agencies and approved for publication. Dr. Limin Yang's work was performed under US Geological Survey contract 1434-CR-97-CN-40274. Dr. Steven Stehman's work was performed under US Environmental Protection Agency contract 0V1036NTSX, and contract 99-9011-P1114C with Raytheon. The authors would like to thank the reviewers for their helpful comments. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use.

References

Anderson, J. F., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. In: *U.S. Geological Survey Professional Paper*. 964 (pp.28). Washington, DC: U.S. Geological Survey.

Congalton, R., & Green, K. (1993). A practical look at sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing, 59* (5), 641–644.

Fegeas, R. G., Claire, R. W., Guptill, S. C., Anderson, K. E., & Hallam, C. A. (1983). Land use and land cover digital data. In: *U.S. Geological Survey Circular* 895-E (pp. 21). Washington, DC: U.S. Geological Survey.

Gopal, S., & Woodcock, C. (1994). Theory and methods for accuracy assessment of thematic maps using fuzzy sets. *Photogrammetric Engineering and Remote Sensing, 60* (2), 181–188.

Hammond, T. O., & Verbyla, D. L. (1996). Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing, 17*, 1261–1266.

Kelly, P. M., & White, J. M. (1993). Preprocessing remotely sensed data for efficient analysis and classification. Applications of Artificial Intelligence 1993: Knowledge-Based Systems in Aerospace and Industry. *Proceedings of SPIE, 1963*, 24–30.

Khorram, S., Biging, G., Colby, D., Congalton, R., Dobson, J., Ferguson, R., Goodchild, M., Jensen, J., & Mace, T. (1999). *Accuracy assessment of remote sensing derived change products. Monograph*. Bethesda, MD: American Society of Photogrammetry and Remote Sensing (64 pp.).

Loveland, T. R., & Shaw, D. M. (1996). Multiresolution land characterization: building collaborative partnerships. In: J. M. Scott, T. Tear, & F. Davis (Eds.), *Gap analysis: a landscape approach to biodiversity planning*. Proceedings of the ASPRS/GAP Symposium, Charlotte, North Carolina. (pp. 83–89). Moscow, ID: National Biological Service.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of the Environment, 64*, 331–344.

Verbyla, D. L., & Hammond, T. O. (1995). Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing, 16*, 581–587.

Vogelmann, J. E., Seevers, P. M., & Oimoen, M. (1997). *Effects of selected variables for discriminating land cover: multiseasonal data, different clustering algorithms, and varying numbers of clusters*. Proceedings of the Pecora 13 Symposium, Sioux Falls, SD, August 1996. (CD-ROM).

Vogelmann, J. E., Sohl, T., Campbell, P. V., & Shaw, D. M. (1998). Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. *Environmental Monitoring and Assessment, 51*, 415–428.

Vogelmann, J. E., Sohl, T., & Howard, S. M. (1998). Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering and Remote Sensing, 64* (1), 45–57.

Yang, L., Stehman, S. V., Wickham, J. D., Smith, J. H., & Van Driel, N. J. (2000). *Thematic validation of land cover data of the eastern United States using aerial photography: feasibility and challenges*. In: G. B. M. Heuvenlink & M. J. P. M. Lemmens (Eds.), Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. (pp. 747–754). The Netherlands: Delft University Press.

Zhu, Z., Yang, L., Stehman, S. V., & Czaplewski, R. L. (1999). Designing an accuracy assessment for a USGS regional land cover mapping program. In: K. Lowell, & A. Jaton (Eds.), *Spatial accuracy assessment — land information uncertainty in natural resources* (pp. 393–398). Chelsea, MI: Sleeping Bear Press/Ann Arbor Press.

Zhu, Z., Yang, L., Stehman, S. V., & Czaplewski, R. L. (2000). Accuracy assessment for the U.S. Geological Survey regional land cover mapping program: New York and New Jersey Region. *Photogrammetric Engineering and Remote Sensing, 66* (12), 1425–1435.